# BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

## IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.**

(19) JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: 09081185 A

(43) Date of publication of application: 28 . 03 . 97

(51) Int. Cl G10L 3/00
G10L 3/00

(21) Application number: 07234043

(22) Date of filing: 12 . 09 . 95

(71) Applicant: ATR ONSEI HONYAKU TSUSHIN KENKYUSHO:KK

(72) Inventor: SHIMIZU TORU
MATSUNAGA SHOICHI
KOSAKA YOSHINORI

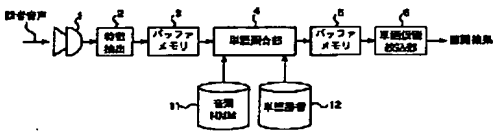
(54) CONTINUOUS VOICE RECOGNITION DEVICE

(57) Abstract:

PROBLEM TO BE SOLVED: To provide a continuous voice recognition device which performs continuous voice recognition of natural uttering with a smaller computing cost than the one using a conventional method.

SOLUTION: A word collating section 4 detects a word-hypothesis of an uttered voice sentence based on the feature parameters of the voice signals of the inputted uttered voice sentence using a one pass.viterbi decoding method, for example, computes the liklihood and outputs them. A word narrowing-down section 6 performs word-hypothesis narrowing-down so that the word-hypotheses of the same word outputted from the section 4 through a buffer memory 5 having the same completion time and a different starting time are represented by a single word-hypothesis having a highest liklihood among the entire liklihood computed from the starting time of uttering to the completion time of the work for every leading phoneme environment of the word.





(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号

**特開平9－81185**

(43)公開日　平成９年(1997)３月28日

| (51)Int.Cl.[6] | 識別記号 | 庁内整理番号 | ＦＩ | 技術表示箇所 |
|---|---|---|---|---|
| Ｇ１０Ｌ　3/00 | ５６１ | | Ｇ１０Ｌ　3/00 | ５６１Ｇ |
| | ５３５ | | | ５３５ |

審査請求　有　　請求項の数１　ＯＬ　（全　６　頁）

(21)出願番号　特願平7－234043

(22)出願日　平成７年(1995)９月12日

(71)出願人　593118597
株式会社エイ・ティ・アール音声翻訳通信研究所
京都府相楽郡精華町大字乾谷小字三平谷５番地

(72)発明者　清水　徹
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内

(74)代理人　弁理士　青山　葆　（外２名）

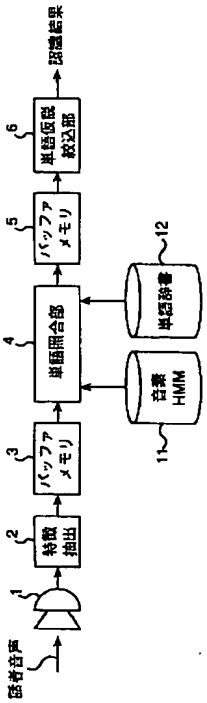
最終頁に続く

(54)【発明の名称】　連続音声認識装置

(57)【要約】

【課題】　従来例に比較してより小さい計算コストで自然発話の連続音声認識を行うことができる連続音声認識装置を提供する。

【解決手段】　単語照合部４は、例えばワンーパス・ビタビ復号化法を用いて、入力される発声音声文の音声信号の特徴パラメータに基づいて上記発声音声文の単語仮説を検出し尤度を計算して出力する。単語絞込部６は、単語照合部４からバッファメモリ５を介して出力される、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する１つの単語仮説で代表させるように単語仮説の絞り込みを行う。



1

【特許請求の範囲】

【請求項１】　入力される発声音声文の音声信号に基づいて上記発声音声文の単語仮説を検出し尤度を計算することにより、連続的に音声認識する音声認識手段を備えた連続音声認識装置において、

上記音声認識手段は、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する１つの単語仮説で代表させるように単語仮説の絞り込みを行うことを特徴とする連続音声認識装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、入力される発声音声文の音声信号に基づいて連続的に音声認識する連続音声認識装置に関する。

【０００２】

【従来の技術】従来から、本特許出願人は、自然発話の音声認識を目的として、連続音声認識系（以下、第１の従来例という。）の開発を進めている（例えば、従来文献１「Ｎａｇａｉ，Ｔａｋａｍｉ，Ｓａｇａｙａｍａ，“Ｔｈｅ　ＳＳＳ－ＬＲ　Ｃｏｎｔｉｎｕｏｕｓ　Ｓｐｅｅｃｈ　Ｒｅｃｏｇｎｉｔｉｏｎ　Ｓｙｓｔｅｍ：Ｉｎｔｅｇｒａｔｉｎｇ　ＳＳＳ－Ｄｅｒｉｖｒｄ　Ａｌｌｏｐｏｈｎｅ　Ｍｏｄｅｌｓ　ａｎｄ　ａ　Ｐｈｏｎｅｍｅ－Ｃｏｎｔｅｘｔ－ＤｅｐｅｎｄｅｎｔＬＲ　Ｐａｒｓｅｒ”，Ｐｒｏｃ．ｏｆ　ＩＣＳＬＰ９２，ｐｐ．１５１１－１５１４，１９９２年」及び従来文献２「Ｓｈｉｍｉｚｕ，Ｍｏｎｚｅｎ，Ｓｉｎｇｅｒ，Ｍａｔｓｕｎａｇａ，“Ｔｉｍｅ－Ｓｙｎｃｈｒｏｎｏｕｓ　Ｃｏｎｔｉｎｕｏｕｓ　Ｓｐｅｅｃｈ　Ｒｅｃｏｇｎｉｚｅｒ　Ｄｒｉｖｅｎ　ｂｙａ　Ｃｏｎｔｅｘｔ－Ｆｒｅｅ　Ｇｒａｍｍａｒ”，Ｐｒｏｃ．ｏｆ　ＩＣＡＳＳＰ９５，ｐｐ．５８４－５８７，１９９５年」参照。）。この第１の従来例では、入力される発生音声文の音声信号に基づいて、音素隠れマルコフモデル（以下、隠れマルコフモデルをＨＭＭという。）と単語辞書を用いて、発声開始からの単語の履歴及び文法状態を管理しながら、音声認識を行っている。

【０００３】一方、単語グラフを用いた音声認識方法（以下、第２の従来例という。）が、従来文献３「Ｎｅｙ，Ａｕｂｅｒｔ，“Ａ　Ｗｏｒｄ　Ｇｒａｐｈ　Ａｌｇｏｒｉｔｈｍ　ｆｏｒ　Ｌａｒｇｅ　Ｖｏｃａｂｕｌａｒｙ，　Ｃｏｎｔｉｎｕｏｕｓ　Ｓｐｅｅｃｈ　Ｒｅｃｏｇｎｉｔｉｏｎ”，Ｐｒｏｃ．ｏｆ　ＩＣＳＬＰ９４，ｐｐ．１３５５－１３５８，１９９４年」及び従来文献４「Ｗｏｏｄｌａｎｄ，Ｌｅｇｇｅｔｔｅｒ，Ｏｄｅｌｌ，Ｖａｌｔｃｈｅｖ，Ｙｏｕｎｇ，“Ｔｈｅ　１９９４　ＨＴＫ　Ｌａｒｇｅ　Ｖｏｃａｂｕｌａｒｙ　Ｓｐｅｅｃｈ　Ｒｅｃｏｇｎｉｔｉｏｎ　Ｓｙｓｔｅｍ”，Ｐｒｏｃ．ｏｆ　ＩＣＡＳＳＰ９５，ｐｐ．７３－７６，１９９５年」において提案されている。

2

【０００４】この第２の従来例の単語グラフの主たるアイデアは、音声認識におけるあいまいさが比較的高い音声信号の領域において単語仮説の候補を処理するということである。この利点は、純粋の音声認識は言語モデルのアプリケーションとは切り離されていることと、複雑な言語モデルは、現在認識中の単語に続く公知のステップに適用することができることである。単語仮説の候補の数は音声認識におけるあいまいさのレベルに対応して変化する必要がある。良い単語グラフを効率的に構築するときの困難さは次の通りである。単語の開始時刻は、一般的に、先行する単語に依存している。第１の近似においては、この依存性を直前の先行単語に対して制限を加えることにより、以下に示すようないわゆる単語ペア近似法を得ている。すなわち、単語のペアとその終了時刻が与えられたときに、２つの単語の間の単語境界は別の先行する単語に独立であるということである。この単語ペア近似法は、本来、複数の文又はｎ個のベスト（最良）である文を効率的に計算するために導入されてきた。この単語グラフは、ｎ個のベストを得るアプローチの方法（以下、ｎベスト法という。）よりも効率的であると期待されている。この単語グラフを用いた方法では、複数の単語仮説を局所的にのみ発生する必要がある一方、ｎベスト法においては、各局所的な単語仮説の候補は、ｎ個のベストである文のリストに対して加えるべき全体の文を必要としている。

【０００５】

【発明が解決しようとする課題】しかしながら、第１の従来例においては、発声開始からの単語の履歴及び文法状態を管理する必要があるため、間投詞の挿入や、言い淀み、言い直しが頻繁に生じる自然発話の認識に用いた場合、単語仮説の併合又は分割に要する計算コストが極めて大きいという問題点があった。すなわち、音声認識のために必要な処理量が大きくなって比較的大きな記憶容量を有する記憶装置が必要となる一方、処理量が大きくなるので処理時間が長くなるという問題点があった。

【０００６】また、上記第２の従来例の単語ペア近似法においては、先行単語毎に１つの仮説で代表させるが、いまだ近似効果は比較的小さい。このため、上記第１の従来例と同様の問題点が生じる。

【０００７】本発明の目的は以上の問題点を解決し、従来例に比較してより小さい計算コストで自然発話の連続音声認識を行うことができる連続音声認識装置を提供することにある。

【０００８】

【課題を解決するための手段】本発明に係る連続音声認識装置は、入力される発声音声文の音声信号に基づいて上記発声音声文の単語仮説を検出し尤度を計算することにより、連続的に音声認識する音声認識手段を備えた連

続音声認識装置において、上記音声認識手段は、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する１つの単語仮説で代表させるように単語仮説の絞り込みを行うことを特徴とする。

【０００９】

【発明の実施の形態】以下、図面を参照して本発明に係る実施形態について説明する。図１に本発明に係る一実施形態の連続音声認識装置のブロック図を示す。本実施形態の連続音声認識装置は、公知のワンーパス・ビタビ復号化法を用いて、入力される発声音声文の音声信号の特徴パラメータに基づいて上記発声音声文の単語仮説を検出し尤度を計算して出力する単語照合部４を備えた連続音声認識装置において、単語照合部４からバッファメモリ５を介して出力される、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する１つの単語仮説で代表させるように単語仮説の絞り込みを行う単語仮説絞込部６を備えたことを特徴とする。

【００１０】図１において、単語照合部４に接続され、例えばハードディスクメモリに格納される音素ＨＭＭ１１は、各状態を含んで表され、各状態はそれぞれ以下の情報を有する。

（ａ）状態番号

（ｂ）受理可能なコンテキストクラス

（ｃ）先行状態、及び後続状態のリスト

（ｄ）出力確率密度分布のパラメータ

（ｅ）自己遷移確率及び後続状態への遷移確率

なお、本実施例において用いる音素ＨＭＭは、各分布がどの話者に由来するかを特定する必要があるため、所定の話者混合ＨＭＭを変換して作成する。ここで、出力確率密度関数は３４次元の対角共分散行列をもつ混合ガウス分布である。

【００１１】また、単語照合部４に接続され、例えばハードディスクに格納される単語辞書１２は、音素ＨＭＭ１１の各単語毎にシンボルで表した読みを示すシンボル列を格納する。

【００１２】図１において、話者の発声音声はマイクロホン１に入力されて音声信号に変換された後、特徴抽出部２に入力される。特徴抽出部２は、入力された音声信号をＡ／Ｄ変換した後、例えばＬＰＣ分析を実行し、対数パワー、１６次ケプストラム係数、Δ対数パワー及び１６次Δケプストラム係数を含む３４次元の特徴パラメータを抽出する。抽出された特徴パラメータの時系列はバッファメモリ３を介して単語照合部４に入力される。

【００１３】単語照合部４は、ワンーパス・ビタビ復号化法を用いて、バッファメモリ３を介して入力される特徴パラメータのデータに基づいて、音素ＨＭＭ１１と単語辞書１２とを用いて単語仮説を検出し尤度を計算して出力する。ここで、単語照合部４は、各時刻の各ＨＭＭの状態毎に、単語内の尤度と発声開始からの尤度を計算する。尤度は、単語の識別番号、単語の開始時刻、先行単語の違い毎に個別にもつ。また、計算処理量の削減のために、音素ＨＭＭ１１及び単語辞書１２とに基づいて計算される総尤度のうちの低い尤度のグリッド仮説を削減する。単語照合部４は、その結果の単語仮説と尤度の情報を発声開始時刻からの時間情報（具体的には、例えばフレーム番号）とともにバッファメモリ５を介して単語仮説絞込部６に出力する。



【００１４】単語仮説絞込部６は、単語照合部４からバッファメモリ５を介して出力される単語仮説に基づいて、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する１つの単語仮説で代表させるように単語仮説の絞り込みを行った後、絞り込み後のすべての単語仮説の単語列のうち、最大の総尤度を有する仮説の単語列を認識結果として出力する。本実施形態においては、好ましくは、処理すべき当該単語の先頭音素環境とは、当該単語より先行する単語仮説の最終音素と、当該単語の単語仮説の最初の２つの音素とを含む３つの音素並びをいう。

【００１５】例えば、図２に示すように、（ｉ－１）番目の単語$W_{i-1}$の次に、音素列$a_1$，$a_2$，…，$a_n$からなるｉ番目の単語$W_i$がくるときに、単語$W_{i-1}$の単語仮説として６つの仮説Ｗａ，Ｗｂ，Ｗｃ，Ｗｄ，Ｗｅ，Ｗｆが存在している。ここで、前者３つの単語仮説Ｗａ，Ｗｂ，Ｗｃの最終音素は／ｘ／であるとし、後者３つの単語仮説Ｗｄ，Ｗｅ，Ｗｆの最終音素は／ｙ／であるとする。終了時刻$t_e$と先頭音素環境が等しい仮説（図２では先頭音素環境が“ｘ／$a_1$／$a_2$”である上から３つの単語仮説）のうち総尤度が最も高い仮説（例えば、図２において１番上の仮説）以外を削除する。なお、上から４番めの仮説は先頭音素環境が違うため、すなわち、先行する単語仮説の最終音素がｘではなくｙであるので、上から４番めの仮説を削除しない。すなわち、先行する単語仮説の最終音素毎に１つのみ仮説を残す。図２の例では、最終音素／ｘ／に対して１つの仮説を残し、最終音素／ｙ／に対して１つの仮説を残す。

【００１６】以上の実施形態においては、当該単語の先頭音素環境とは、当該単語より先行する単語仮説の最終音素と、当該単語の単語仮説の最初の２つの音素とを含む３つの音素並びとして定義されているが、本発明はこれに限らず、先行する単語仮説の最終音素と、最終音素と連続する先行する単語仮説の少なくとも１つの音素とを含む先行単語仮説の音素列と、当該単語の単語仮説の最初の音素を含む音素列とを含む音素並びとしてもよい。

5

【００１７】

【実施例】本発明者は、図１の連続音声認識装置の有効性を確認するために、自然発話データベースを用いて単語グラフ生成実験を行なった。“トラベル・プランニング”をタスクとした本出願人が所有する音声言語データベース（例えば、従来文献５「Ｍｏｒｉｍｏｔｏ ｅｔ ａｌ．，“Ａ Ｓｐｅｅｃｈ ａｎｄ Ｌａｎｇｕａｇｅ Ｄａｔａｂａｓｅ ｆｏｒ Ｓｐｅｅｃｈ ＴｒａｎｓｌａｔｉｏｎＲｅｓｅａｒｃｈ”，Ｐｒｏｃ．ｏｆ ＩＣＳＬＰ９４，ｐｐ．１７９１－１７９４，１９９４年」参照。）の「ホテル予約」に関する対話（申込者側５話者の発声：５対話，５６発声，６８７語）を用いて評価した。音響分析は、標本化周波数１２ｋＨｚ，フレーム間隔５ｍｓｅｃ，ハミング窓２０ｍｓｅｃの仕様で分析し、特徴パラメータとして、１～１６次ＬＰＣケプストラム、１～１６次ΔＬＰＣケプストラム、対数パワー、Δ対数パワーを用いた。音響モデル（隠れマルコフ網：４０１状態，５混合）は、朗読音声（１５０文）を用いて学習した音響モデルをさらに上記データベースのテストデータに現れない話者９名の発声（１２８発声）を用いて発話様式に適応した。また、言語モデルは、「ホテル予約」を含む“トラベル・プランニング”全般（１８，３１５発声，２２９，１５９語）を用いて学習した。単語パープレキシティは、５５．９であった。単語辞書（１，１１３語）は、評価データの語彙を全て含んでおり、予め登録されていない未知語（未登録語ともいう。）はないものとした。

【００１８】次いで、開始時刻の異なる単語仮説の絞り込み効果について以下に説明する。図３に、絞り込みを行なった場合（本実施形態）と絞り込みを行なわない場合の各単語仮説の先行単語数の分布の比較を示す。絞り込みを行なうことによって、平均先行単語数が３．５９から１．７０に削減された。また、絞り込みを行なわなかった場合に対して、開始時刻の違いを無視した平均先行単語数を計算したところ、１．３６であった。この結果から、単語の先頭音素環境ごとに１つの仮説で代表させる本発明の方法は、少ない計算量で、先行単語毎に１つの仮説で代表させる第２の従来例の単語ペア近似法にかなり近い効果が得られると考えられる。

【００１９】以上説明したように、本実施形態によれば、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する１つの単語仮説で代表させるように単語仮説の絞り込みを行う。すなわち、先行単語毎に１つの単語仮説で代表させる第２の従来例の単語ペア近似法に比較して、単語の先頭音素の先行音素（つまり、先行単語の最終音素）が等しいものをひとまとめに扱うために、単語仮説数を削減することができ、近似効果は大きい。特に、語彙数が増加した場合において削減効果が大きい。従って、当該連続音声認識装置を、間投詞の挿入や、言い淀み、言い直しが頻繁に生じる自然発話の認識に用いた場合であっても、単語仮説の併合又は分割に要する計算コストは従来例に比較して小さくなる。すなわち、音声認識のために必要な処理量が小さくなり、それ故、単語照合部４のワーキングメモリ（図示せず。）、バッファメモリ５及び単語仮説絞込部６のワーキングメモリ（図示せず。）などの音声認識のための記憶装置において必要な記憶容量は小さくなる一方、処理量が小さくなるので音声認識のための処理時間を短縮することができる。

6

【００２０】

【発明の効果】以上詳述したように本発明によれば、入力される発声音声文の音声信号に基づいて上記発声音声文の単語仮説を検出し尤度を計算することにより、連続的に音声認識する音声認識手段を備えた連続音声認識装置において、上記音声認識手段は、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する１つの単語仮説で代表させるように絞り込みを行う。すなわち、先行単語毎に１つの単語仮説で代表させる第２の従来例の単語ペア近似法に比較して、単語の先頭音素の先行音素（つまり、先行単語の最終音素）が等しいものをひとまとめに扱うために、単語仮説数を削減することができ、近似効果は大きい。特に、語彙数が増加した場合において削減効果が大きい。従って、当該連続音声認識装置を、間投詞の挿入や、言い淀み、言い直しが頻繁に生じる自然発話の認識に用いた場合であっても、単語仮説の併合又は分割に要する計算コストは従来例に比較して小さくなる。すなわち、音声認識のために必要な処理量が小さくなり、それ故、音声認識のための記憶装置において必要な記憶容量は小さくなる一方、処理量が小さくなるので音声認識のための処理時間を短縮することができる。

【図面の簡単な説明】

【図１】　本発明に係る一実施形態である連続音声認識装置のブロック図である。

【図２】　図１の連続音声認識装置における単語仮説絞込部６の処理を示すタイミングチャートである。

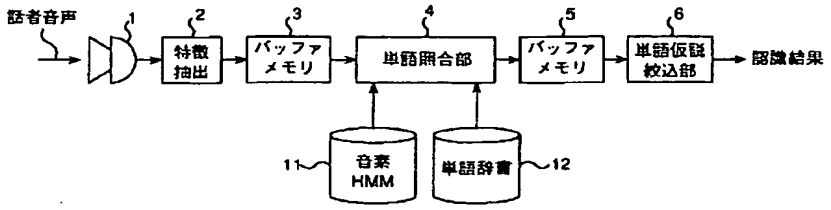
【図３】　図１の連続音声認識装置の実験結果における、単語間の遷移における単語仮説の絞り込み効果を示す先行単語の個数に対するノード数のグラフである。

【符号の説明】

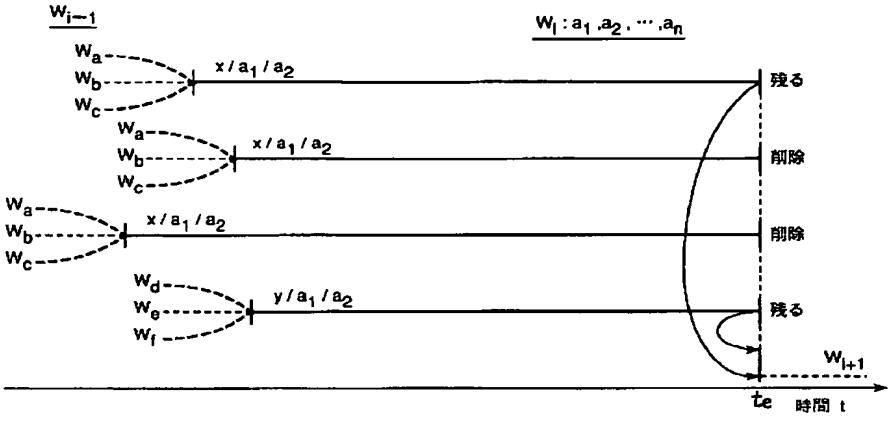
１…マイクロホン、
２…特徴抽出部、
３，５…バッファメモリ、
４…単語照合部、
６…単語仮説絞込部、
１１…音素ＨＭＭ、

１２…単語辞書。

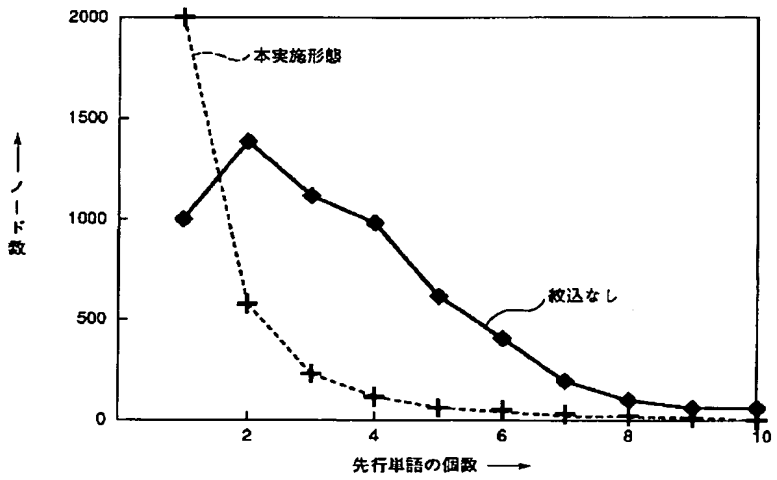
【図１】




【図２】




【図３】



フロントページの続き

(72)発明者　松永　昭一
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内

(72)発明者　匂坂　芳典
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内